Estudio de herramientas para la gestión de repositorios digitales

Historia de las revisiones

Fecha	**Versión**	**Descripción**	**Autor**
2011/07/06	1.0	Creación del Documento	Alexander Barón Salazar
2011/09/25	2.0	Corregido	Alexander Barón Salazar

CONTENIDO

Pág.

Lista de Figuras

Pág. 1

1. INTRODUCCION

A fin de identificar la herramienta tecnológica más apropiada para apoyar la gestión de los activos de software, se realizó un estudio comparativo de tres de las herramientas de gestión de repositorios digitales más reconocidas por la comunidad usuaria: DSpace (FEDORA, 2011) E-PRINTS (UNIVERSITY OF SOUTHAMPTON, 2011) y FEDORA COMMONS (FEDORA, 2005).

A continuación se exponen los criterios de análisis, el resumen del estudio y se presenta la herramienta seleccionada.

2. CRITERIOS DE ANÁLISIS

2.1 Descripción General

Es un resumen de la presentación de la herramienta. Presenta aspectos relacionados con la empresa propietaria, tipo de software, sus desarrolladores y una breve reseña histórica. Este criterio provee al lector una idea general de la herramienta objeto de análisis.

2.2 Requisitos de software

Describe los requisitos de software mínimos necesarios para el correcto funcionamiento de la herramienta. Este criterio permite determinar el nivel de exigencia de la herramienta en aspectos de potencia de hardware y software de plataforma.

2.3 Facilidad de Personalización

Describe las facilidades que provee la herramienta para personalizar y adaptar las funcionalidades a los requisitos particulares de un repositorio digital específico accediendo al código, utilizando plantillas o herramientas de parametrización. Este criterio permite determinar la flexibilidad de la herramienta para adaptarla a las condiciones específicas de un repositorio digital.

2.4 Documentación disponible

Hace referencia a las fuentes de información existentes y de fácil acceso para soportar la configuración, adaptación, uso y mantenimiento del repositorio digital. Este criterio permite determinar el nivel de respaldo que tiene la herramienta cuando el usuario requiere de ayuda para la implementación de cualquier tarea o funcionalidad en el proceso de construcción o uso de un repositorio digital.

2.5 Comunicación con el Usuario

Describe las funcionalidades que provee la herramienta para adaptar la interfaz de usuario tanto de entrada como de salida a las características particulares del tipo de recurso digital. Este criterio permite establecer el nivel de flexibilidad de adaptación de las interfaces.

2.6 Mecanismos de Descripción de recursos digitales

Describe el mecanismo empleado por la herramienta para la descripción de los recursos digitales del repositorio. Este criterio permite establecer la flexibilidad de

la herramienta para configurar los metadatos a las características de un recurso digital específico.

2.7 Estadísticas

Describe las funcionalidades que provee la herramienta para proporcionar información que permita el análisis estadístico sobre diferentes aspectos de los recursos digitales del repositorio.

2.8 Gestión de Usuarios

Describe las funcionalidades que provee la herramienta para la gestión de usuarios, el mecanismo de control y la asignación de permisos y niveles de acceso. Este criterio permite establecer la seguridad con que la herramienta permite gestionar los recursos digitales del repositorio.

2.9 Gestión de los recursos digitales

Describe los mecanismos para el desarrollo de las funcionalidades básicas que operan sobre los recursos del repositorio digital tales como: adición, eliminación, modificación.

2.10 Mecanismos de búsqueda

Describe el mecanismo empleado para ubicar un recurso digital específico en el repositorio a partir de criterios de búsqueda pre-establecidos. Este criterio permite establecer la flexibilidad de la herramienta para determinar criterios de búsqueda acordes con las características y mecanismo de descripción de un tipo de recurso digital.

2.11 Mecanismos de Recuperación

Describe la forma como la herramienta muestra y deja a disposición del usuario un recurso digital que previamente ha sido buscado. Este criterio permite establecer la flexibilidad de la herramienta para adaptar el mecanismo de visualización del recurso digital de acuerdo con sus características particulares y el tipo de archivo que lo contiene.

2.12 Gestión de experiencias

Describe las funcionalidades que provee la herramienta para gestionar las experiencias de uso de los recursos digitales del repositorio. Este criterio permite determinar la eficiencia de la herramienta para gestionar las impresiones de los

usuarios acerca del recurso digital utilizado, información que puede ser consultada por otros usuarios para establecer el nivel de aceptación del recurso.

2.13 Mecanismos de almacenamiento

Describe el proceso y la estructura de datos empleada por la herramienta para almacenar los recursos digitales del repositorio. Este criterio permite determinar la flexibilidad de la herramienta para adaptar el repositorio a las características particulares del recurso digital gestionado y el tipo de repositorio que se pretende construir.

3. DSPACE 1.6.2

A continuación se aplican los criterios de análisis a la herramienta DSpace versión 1.6.2.

3.1 Descripción General

DSpace es una herramienta para la gestión de repositorios digitales gratuita y de libre acceso que provee un conjunto de herramientas para la administración de colecciones digitales. DSpace fue desarrollado por Hewlet Packard (HP COMPANY, 2011) y el MIT (MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 2011) y la primera versión se liberó en el año 2002.

3.2 Requisitos de Software

Los requisitos de software para el correcto funcionamiento de DSpace son (FEDORA, 2011):

- Sistema Operativo: Unix o similar (hp/ox, Solaris, Linux, etc.). Sistema operativo Windows XP en adelante.
- Máquina Virtual: Sun Java 5 o mayor.
- Compilador: Apache Maven 2.0.8 o mayor.
- Compilador y constructor: Apache Ant 1.7 o mayor.
- Base de Datos Relacional: PostgresSQL 7.3 o mayor u Oracle 9 o mayor.
- Cualquiera de las siguientes opciones para Motor Servlet (servidor): Jakarta Tomcat 4 o mayor, Jetty, Caucho, Resin o software equivalente.

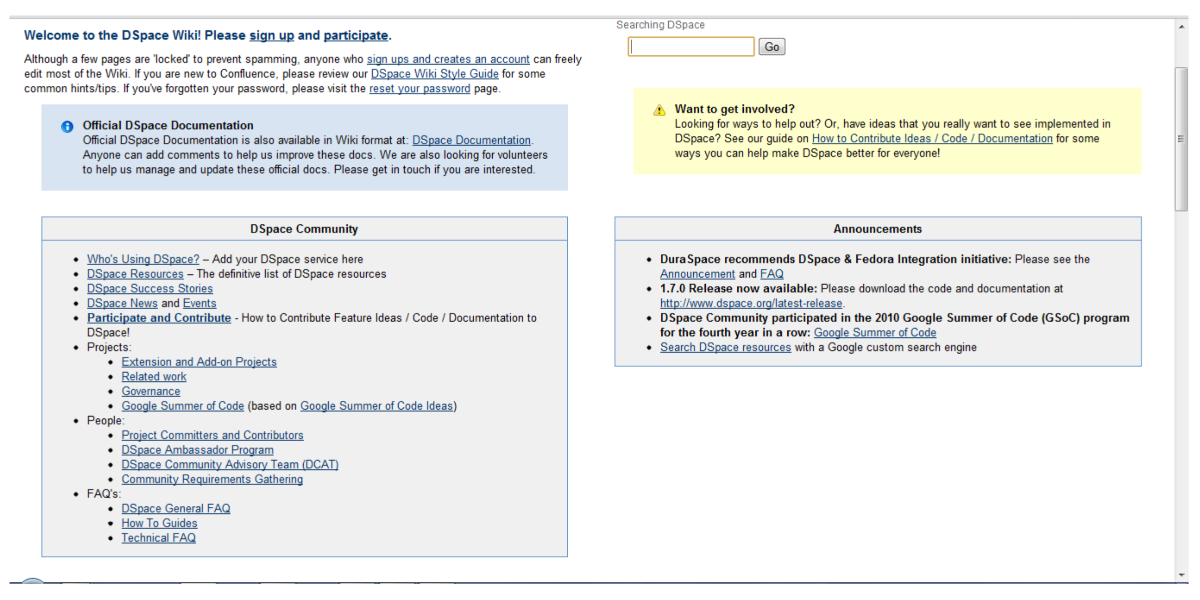
3.3 Facilidad de Personalización

La modularidad de DSpace, permite dos niveles de personalización de forma independiente: básica y avanzada, acorde con las capas de su arquitectura. Estas funcionalidades las provee por medio del manejo de archivos de configuración jsp (JavaServer Pages), css (Cascading Style Sheets), xhtml (eXtensible Hypertext Markup Language), xls(Extension para Libros Contables de Microsoft), xml(eXtensible Markup Language) (FEDORA, 2011).

3.4 Documentación Disponible

DSpace ofrece variedad de información para su configuración, adaptación y uso, incluye un video que expone las características básicas del producto, un detallado manual de usuario donde describe su estructura, procesos de instalación, mecanismos de configuración personalizada, y preguntas frecuentes (FEDORA, 2011).

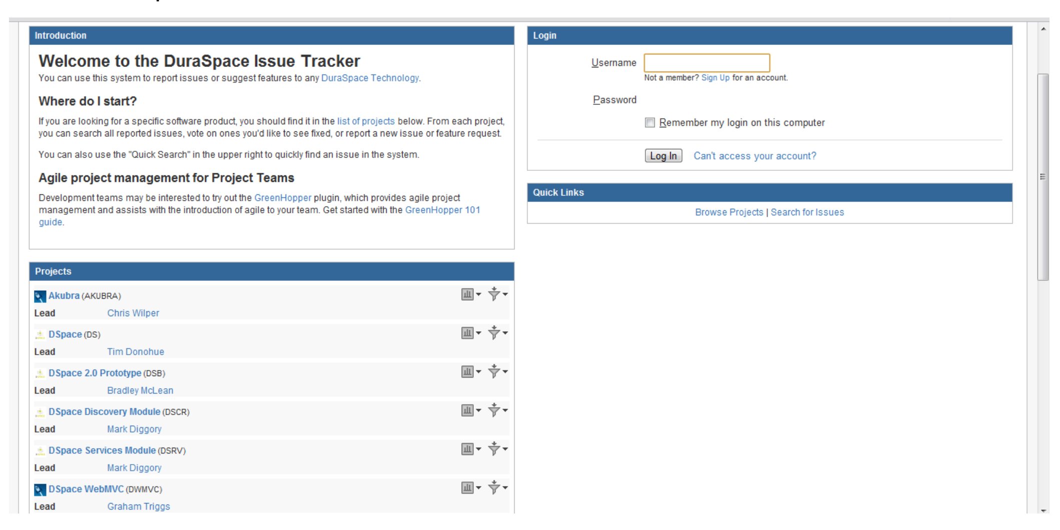
Figura 1. Wiki DSpace Documentación



Fuente: (FEDORA, 2011)

Actualmente existe la comunidad de usuarios de DSpace que integra un conjunto de servicios como: comunidad de correos, foros de discusión, embajadores regionales DSpace, blogs Duraspace, comunidad de programadores (FEDORA, 2011) y Jira (FEDORA, 2011). Jira es una plataforma que monitorea el trabajo de desarrollo colaborativo y realiza los pedidos de características por parte de los usuarios además permite a la comunidad observar problemas específicos, recibir actualizaciones cuando haya cambios y priorizar problemas utilizando un sistema de votación.

Figura 2. JIRA DSpace



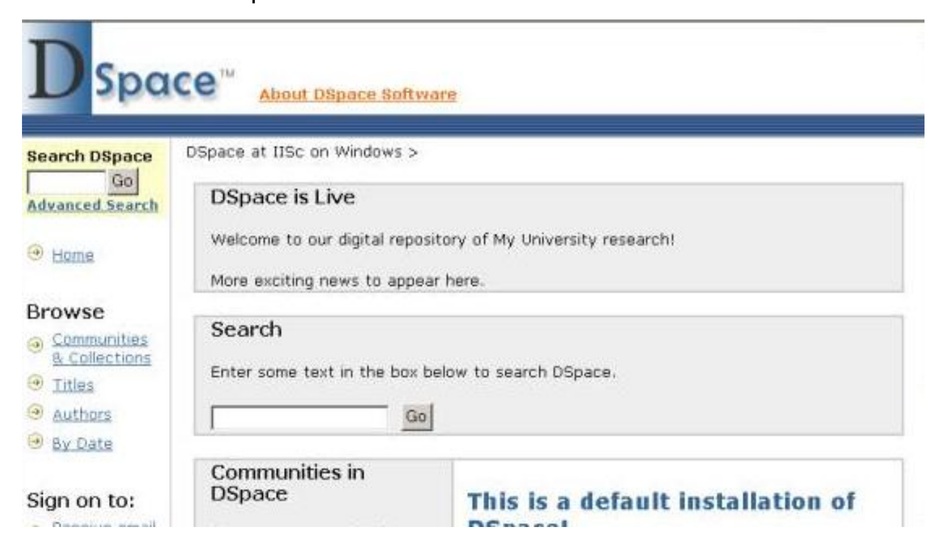
Fuente: (FEDORA, 2011)

3.5 Comunicación con el Usuario

DSpace integra los elementos de visualización en la capa de aplicación que provee el componente administrador Web UI (Interfaz de Usuario Web). DSpace dispone de dos tipos de interfaz:

- Jspui (JavaServer Pages) que permite una rápida creación de contenido web dinámico. La interfaz Jspui está compuesta por Java Servlets que manejan la lógica de negocio y por JavaServer Pages que producen las paginas html (FEDORA, 2011).

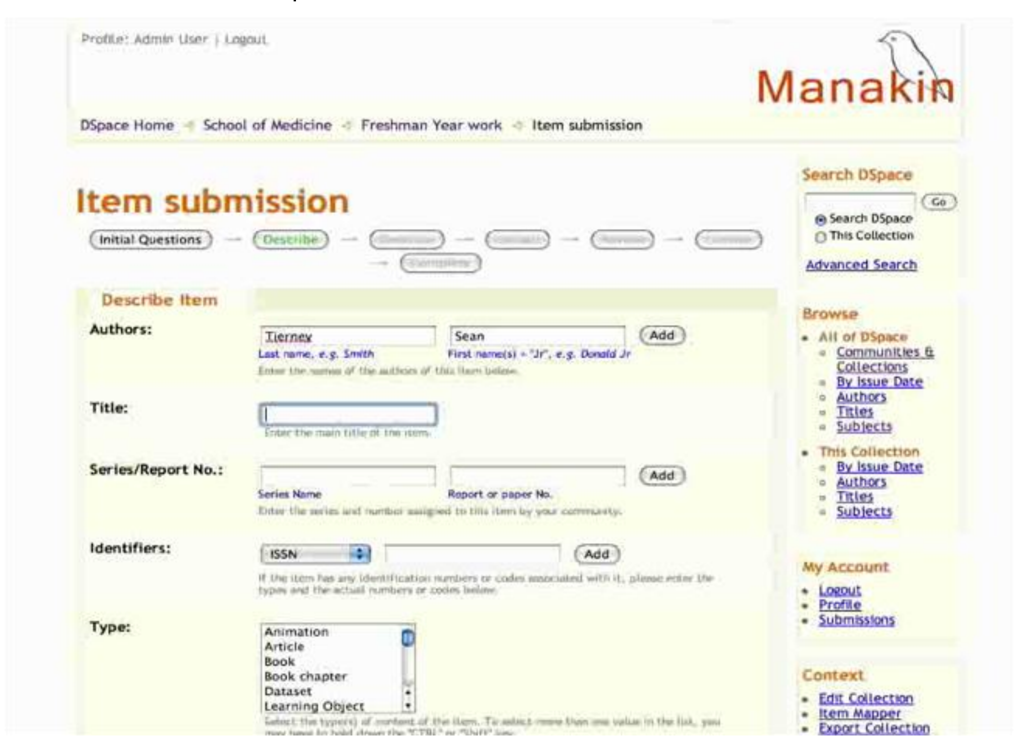
Figura 3. Interfaz JSPUI de DSpace



Fuente: (FEDORA, 2011)

- Xmlui (Manakin Framework), basado en el Framework de Apache Cocoon es una interfaz de usuario modular que se puede personalizar a nivel del repositorio, comunidad o colección. Esta interfaz soporta la internacionalización de DSpace. Xmlui permite altos niveles de adaptabilidad debido a que separa la lógica de negocio de la lógica de presentación (FEDORA, 2011).

Figura 4. Interfaz XMLUI de DSpace



Fuente: (FEDORA, 2011)

3.6 Mecanismos de Descripción de Recursos Digitales (FEDORA, 2011)

DSpace soporta tres tipos de metadatos para la descripción de recursos digitales:

- Metadatos Descriptivos: DSpace soporta múltiples esquemas de metadatos planos para descripción de recursos digitales. Por defecto, DSpace provee el esquema de metadatos Qualified Dublin Core, un grupo de elementos y calificadores utilizados por la biblioteca del MIT que vienen pre-configurados. DSpace provee las facilidades para adaptar este esquema o crear nuevos de acuerdo con las características de un tipo de recurso digital específico.

- Metadatos Administrativos: Es el conjunto de metadatos para describir los parámetros de gestión del repositorio y las políticas de autorización. Estos metadatos están contenidos en el gestor de base de datos de DSpace.

- Metadatos Estructurales: Son los metadatos que describen la información acerca de cómo presentar un recurso digital a un usuario final y las relaciones entre las partes constituyentes de un elemento del repositorio.

3.7 Estadísticas (FEDORA, 2011)

DSpace ofrece un sistema básico de estadísticas que puede ser remplazado por uno más avanzado instalando un conjunto de plug-ins conocidos como add-ons. El sistema de estadísticas de DSpace proporciona una gran variedad de reportes configurables a nivel de recurso, comunidad y colección. Los reportes estadísticos son generados automáticamente por el sistema utilizando los archivos log de DSpace.

Los reportes pueden ser programados por periodos de tiempo o a nivel general, igualmente, pueden ser restringidos al uso de determinados usuarios, definidos como privados o públicos.

Figura 5: Figura 5. Estadísticas DSpace

Fuente: (FEDORA, 2011)

3.8 Gestión de Usuarios

DSpace clasifica a los usuarios de acuerdo a un nivel de privilegios los cuales son otorgados por el sistema de autorización. Los usuarios que se registran se denominan usuarios anónimos. Los usuarios anónimos únicamente pueden buscar y recuperar ciertos elementos del repositorio. Las e-person (personas registradas en DSpace via web) de las cuales DSpace almacena datos de contacto y de interés general. Las e-person tienen mayor nivel de acceso a los recursos del repositorio e ingresan identificándose con un nombre de usuario y una contraseña (FEDORA, 2011).

Los usuarios se integran en grupos. Los grupos son otro tipo de entidad al que se le asignan privilegios. Explícitamente un grupo es una lista de e-person (grupo de personas registradas) quienes heredan los privilegios otorgados al grupo.

DSpace implementa la autenticación por medio de un mecanismo llamado Autenticación Apilable, que consiste en una pila de métodos de autenticación. Las aplicaciones de DSpace invocan al administrador de autenticación que identifica a la e-person que esta en sesión y el grupo al cual pertenece, de esta manera determina los niveles de acceso en la aplicación que está utilizando (FEDORA, 2011).

Para el correcto funcionamiento de DSpace se requiere de la creación de un usuario administrador principal. El administrador principal es el usuario que tiene acceso y control total sobre el repositorio.

3.9 Gestión de los Recursos Digitales

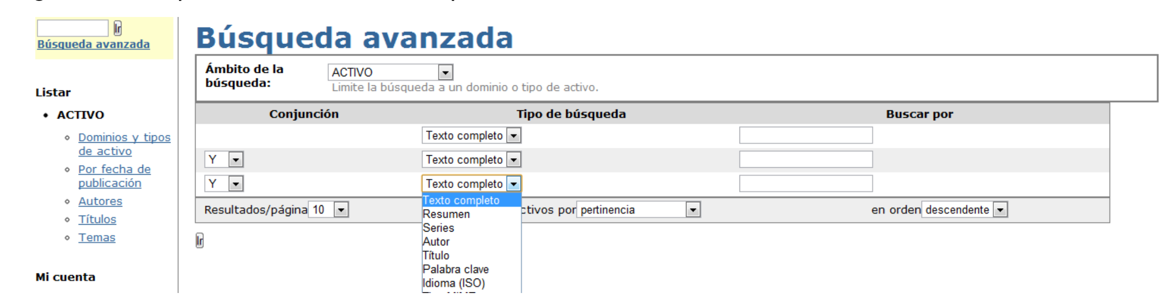
DSpace gestiona funcionalidades básicas sobre los recursos digitales del repositorio: adicionar, eliminar o modificar, por medio de una API de contenido denominada paquete de Gerencia de Contenidos org.DSpace.content. Esta API contiene clases java para la manipulación del contenido almacenado en el sistema DSpace (FEDORA, 2011).

3.10 Mecanismos de Búsqueda

Las expectativas de búsqueda del usuario son muy altas, para satisfacer este requisito, DSpace ofrece múltiples opciones: por medio de una referencia externa o link, buscando una o más palabras claves en los metadatos, por colecciones o por comunidades. La búsqueda e indexación son realizadas utilizando una API que permite indexar nuevo contenido, regenerar índices y realizar búsquedas en toda la comunidad o colección. La búsqueda puede ser avanzada o simple (FEDORA, 2011).

El código de búsqueda de DSpace es una API sencilla que incluye la máquina de búsqueda Lucene. Lucene es un API de código abierto para recuperación de información, originalmente implementada en Java. Lucene esta soportado por el Apache Software Foundation y se distribuye bajo la licencia Apache Software (FEDORA, 2011).

Figura 6. Búsqueda Avanzada de DSpace



Fuente: (FEDORA, 2011)

3.11 Mecanismos de Recuperación

DSpace permite al usuario visualizar todo tipo de recurso digital ya que soporta cualquiera de los formatos conocidos actualmente. DSpace permite almacenar y mostrar contenido HTML siempre y cuando no sea dinámico ni autocontenido. Para algunos formatos, DSpace tiene su propio visor (FEDORA, 2011).

DSpace identifica un punto de referencia de estabilidad del repositorio por medio de una URL. Sin embargo las URL son poco confiables cuando el contenido cambia de ubicación o el sitio web se puede reconfigurar sin ninguna notificación. Para resolver este problema DSpace utiliza un identificador persistente para cada elemento del repositorio. DSpace utiliza el sistema Handle CNRI para crear esos identificadores. Es necesario activar el servidor Handle con el fin de aceptar y resolver las peticiones (FEDORA, 2011).

3.12 Gestión de Experiencias

DSpace provee dos alternativas no excluyentes para gestionar las experiencias de uso de los activos de software. La primera opción es la creación de herramientas como chats y foros para que los usuarios intercambien impresiones sobre el uso de un activo de software en particular o del catálogo en general. La segunda opción para gestionar experiencias en DSpce, es a través de la herramienta relacionador de activos que permite enlazar dos o más activos de software que tengan alguna relación, en este caso la relación se utiliza para indicar que un activo es experiencia de uso de otro activo (FEDORA, 2011).

3.13 Mecanismos de Almacenamiento

DSpace proporciona dos alternativas para el almacenamiento de los recursos digitales: Un sistema de archivos de almacenamiento en servidor y un sistema SRB (Storage Resource Broker). DSpace proporciona un gestor de

almacenamiento sofisticado y robusto con el fin de prestar el servicio de almacenamiento ilimitado y back up del repositorio en otra máquina local o remota (FEDORA, 2011).

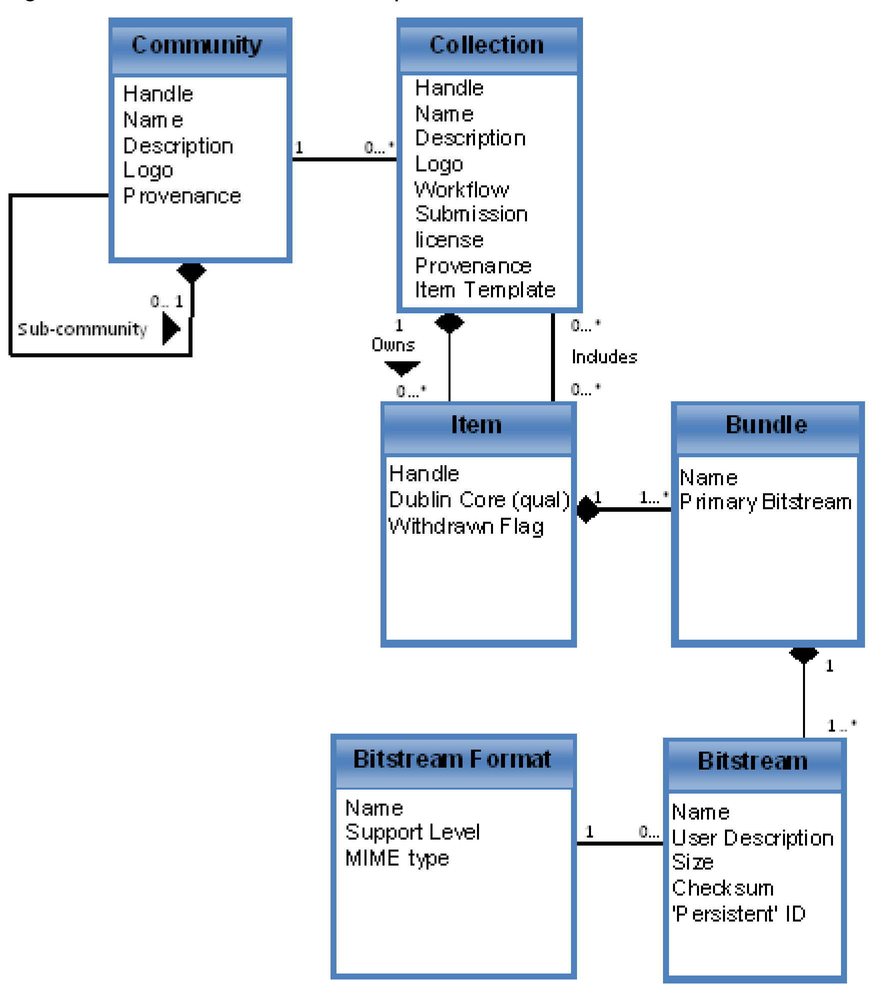
DSpace provee la herramienta Checksum Checker que sin intervención del usuario, verifica el contenido de archivos en diferentes momentos del proceso de registro de un nuevo activo de software, cuando advierte una diferencia, indica una posible alteración no autorizada, envi una notificación al usuario y este puede repetir el proceso. De esta manera, Checksum Checker soporta la preservación de la integridad del contenido del repositorio digital (FEDORA, 2011).

DSpace incorpora los conceptos comunidades y colecciones (Figura 7) para ofrecer una estructura navegable y familiar. Cada servicio de DSpace está compuesto por comunidades, el nivel más alto en la jerarquía de contenido. A su vez cada comunidad está compuesta por colecciones integrada por elementos de contenido (Figura 8).

Figura 7. Estructura General DSpace del Repositorio

Fuente: (FEDORA, 2011)

Figura 8. Estructura de Datos DSpace



Fuente: (FEDORA, 2011)

4. E-PRINTS 3.2.5

4.1 Descripción General

E-Prints (UNIVERSITY OF SOUTHAMPTON, 2011) es un software de código abierto para construir repositorios digitales. E-Prints fue creado en el año 2000 en la Universidad de Southampton (UNIVERSITY OF SOUTHAMPTON, 2011) como resultado de la reunión de Santa Fe realizada en 1999. E-Prints es una de las primeras herramientas desarrolladas para la gestión de repositorios digitales institucionales. Por sus facilidades de uso, rápidamente E-Prints se convirtió en uno de los más utilizados a nivel mundial (UNIVERSITY OF NOTTINGHAM, 2006).Descripción General

4.2 Requisitos de software

Los requisitos de software para el correcto funcionamiento de E-prints son (UNIVERSITY OF SOUTHAMPTON, 2011):

- ✓ Sistema Operativo: Microsoft Windows Xp o mayor.
- ✓ Motor Servlet (servidor): Apache 2.0 o mayor.
- ✓ Base de Datos Relacional :MySQL 5 o mayor.
- ✓ Compilador: ModPerl 2.0 o mayor.

Opcionales:

- ✓ Editor de imagen: ImageMagic
- ✓ Visor de archivos pdf: Xpdf

4.3 Facilidad de Personalización

E-Prints implementa funcionalidades de personalización básica y avanzada (UNIVERSITY OF SOUTHAMPTON, 2011). Los elementos que pueden adaptarse son la interfaz de usuario, la interfaz de búsqueda, la interfaz de recuperación y el conjunto de metadatos de descripción.

E-Prints ofrece varias herramientas para la personalización básica: mediante la edición de archivos de configuración y modificando las variables de parametrización y mediante el uso de plantillas prediseñadas, archivos Perl o APIs.

Para realizar tareas de personalización avanzada se requiere el conocimiento de tecnologías como XHTML, CSS, XML y el lenguaje de programación PERL para el desarrollo de APIs.

4.4 Documentación disponible

E-Prints proporciona una amplia variedad de documentación (UNIVERSITY OF SOUTHAMPTON, 2011) y elementos de soporte para la instalación, configuración y uso de la herramienta. Sin embargo, para su uso eficiente son necesarios conocimientos de programación e informática.

La comunidad de desarrolladores de E-Prints ofrece un soporte adicional en línea.

Los servicios de ayuda al usuario que ofrece E-Prints son:

- Alojamiento y mantenimiento del repositorio
- Personalización del repositorio de acuerdo a las especificaciones del usuario
- Importación de datos
- Soporte técnico en curso
- Entrenamiento personalizado para el personal y administradores del repositorio
- Asesoramiento en política de desarrollo
- Asistencia con abogacía y promoción
- Gerencia de proyecto para el lanzamiento del producto

4.5 Comunicación con el Usuario

E-Prints ofrece una interfaz totalmente reconfigurable (UNIVERSITY OF SOUTHAMPTON, 2011), soportada en tecnologías XHTML y CSS. E-Prints utiliza un archivo tipo plantilla para renderizar todas las páginas con el fin de asegurar su consistencia. La estructura web por defecto incluye:

- Encabezado: zona superior donde puede ir el logotipo o título del repositorio.

- Barra de catalogación: conformada por el inicio de página, información, año, tema, división, autor, supervisores de tesis, tipo de documento, inicio de sesión, crear cuenta, instrucciones.
- Cuerpo de la página: información principal, catalogación, búsqueda básica, búsqueda avanzada, indexadores.
- Pie de página: información adicional.

4.6 Mecanismos de Descripción de recursos digitales

E-Prints utiliza metadatos como mecanismo de descripción de los recursos digitales del repositorio (UNIVERSITY OF SOUTHAMPTON, 2011). Cada metadato se define como un campo al que se le asocia un tipo y una propiedad de nombre predefinidos por E-Prints. Los metadatos son los criterios de indexación y búsqueda de los recursos del repositorio digital.

4.7 Estadísticas

E-Prints provee dos sistemas de estadísticas: uno básico tanto para los recursos digitales a nivel individual como para todo el sistema. Otro avanzado, el Add-On estadístico que provee reportes estadísticos configurables más sofisticados (UNIVERSITY OF SOUTHAMPTON, 2011).

4.8 Gestión de Usuarios

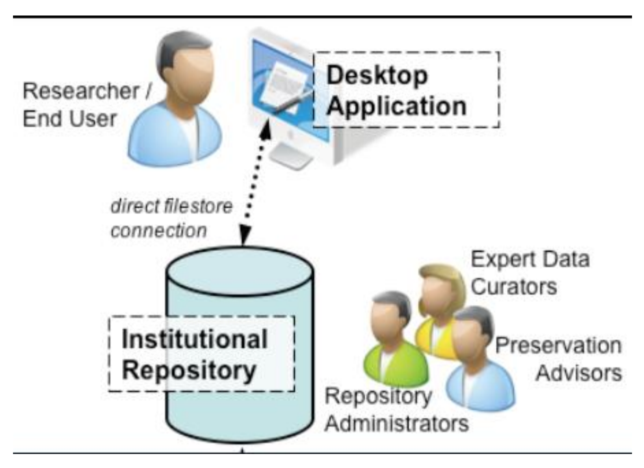
E-Prints gestiona cuatro tipos de usuarios (UNIVERSITY OF SOUTHAMPTON, 2011):

- Investigadores: son usuarios con privilegios especiales que pueden ver y unirse a la comunidad para aportar recursos digitales.
- Administradores del repositorio: son usuarios con privilegios de máximo nivel, le permiten acceder a todas las funcionalidades de la herramienta y realizar tareas de mantenimiento y personalización.
- Experto conservador de datos: es un usuario auxiliar del repositorio que ayuda a mantener la integridad y el orden de los elementos, accede al repositorio con privilegios que le permiten desarrollar tareas puntuales de mantenimiento y operación.
- Consejero de preservación: es un usuario auxiliar del repositorio que ayuda

a que el contenido del material sea propicio para el repositorio, tiene privilegios especiales que le permiten validar los recursos candidatos a ser parte del repositorio digital.

Los usuarios de E-Prints se autentican ante el sistema por medio de un nombre y una contraseña (UNIVERSITY OF SOUTHAMPTON, 2011).

Figura 9. Rol de Usuarios en E-Prints



Fuente: (UNIVERSITY OF SOUTHAMPTON, 2011)

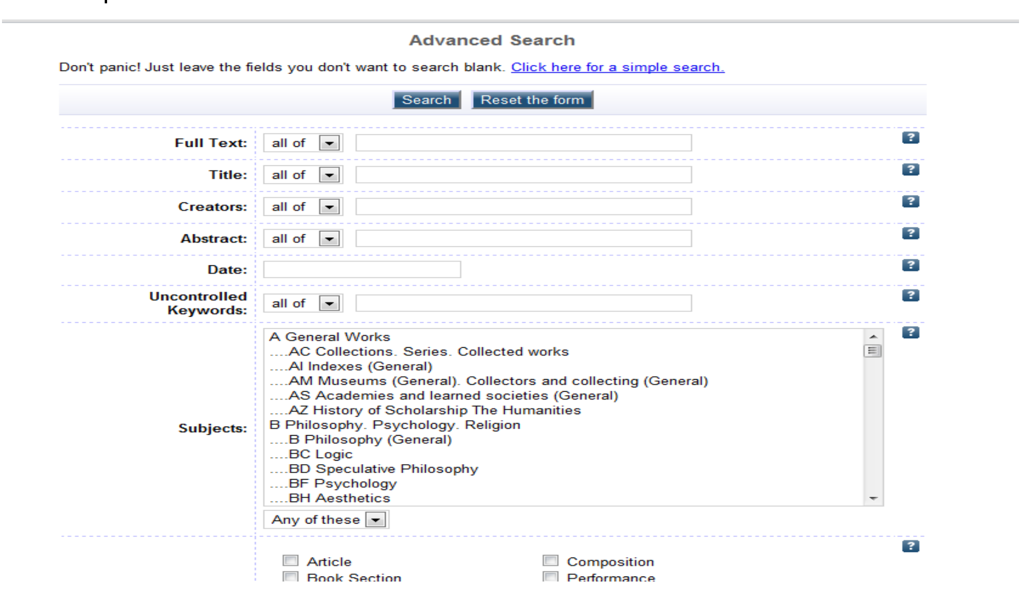
4.9 Gestión de los recursos digitales

E-Prints proporciona herramientas para proveer las funcionalidades básicas del repositorio como adicionar, eliminar y modificar los recursos digitales (UNIVERSITY OF SOUTHAMPTON, 2011). Para realizar las funcionalidades básicas E-prints aplica un conjunto de políticas configurables de preservación de la integridad del repositorio.

4.10 Mecanismos de búsqueda

E-Prints provee un potente sistema de búsqueda simple y avanzada (UNIVERSITY OF SOUTHAMPTON, 2011). Los criterios de búsqueda y presentación de resultados son configurables. La búsqueda simple integra un reducido conjunto de criterios. La búsqueda avanzada se basa en un conjunto más amplio de criterios elegidos por el usuario. Los criterios de búsqueda se basan en el conjunto de metadatos que describen los recursos digitales del repositorio. E-Prints permite al usuario definir el orden de presentación de los resultados de la búsqueda utilizando como criterios de orden los metadatos de descripción.

Figura 10. Búsqueda Avanzada en E-Prints



Fuente: (UNIVERSITY OF SOUTHAMPTON, 2011)

4.11 Mecanismos de Recuperación

E-Prints puede contener la mayoría de los formatos de archivos actuales dentro de su repositorio además de realizar una pre-visualización por medio de thumbnails (imágenes miniatura del contenido) (UNIVERSITY OF SOUTHAMPTON, 2011).

Como valor agregado E-Prints incorpora un sistema de análisis y riesgo de formato de archivo con el fin de ofrecer información estadística de comportamiento.

4.12 Gestión de experiencias

E-Prints provee facilidades para la personalización de la gestión de experiencias por medio de la edición de código para la creación de herramientas como chats y foros. La personalización de la gestión de experiencias en E-Prints depende de la habilidad de los desarrolladores del repositorio digital (UNIVERSITY OF SOUTHAMPTON, 2011).

4.13 Mecanismos de almacenamiento

E-Prints fundamenta los mecanismos de almacenamiento en un conjunto de servicios de gestión de persistencia de los recursos digitales del repositorio (UNIVERSITY OF SOUTHAMPTON, 2011):

- Disco local: los recursos digitales del repositorio pueden ser almacenados a nivel local, es decir en la misma máquina donde se presta el servicio.

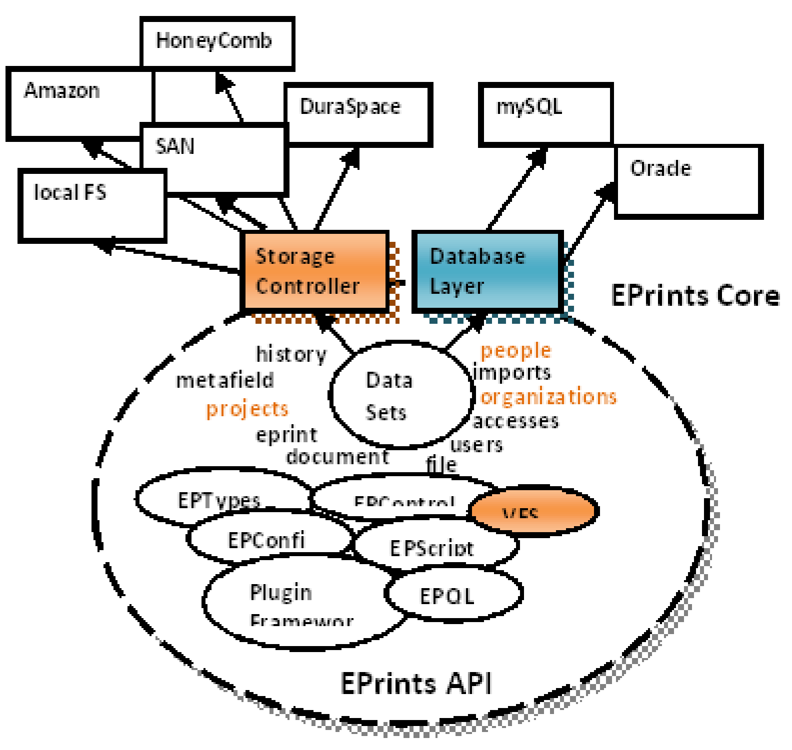
- SAN (Storage Area Network): soporta una red de servidores, arrays de discos y librerías de soporte.

- NAS (Network Attached Storage): soporta tecnologías dedicadas a compartir la capacidad de almacenamiento de un servidor con sus respectivos clientes.

- DuraCloud: provee soporte de preservación y servicios de acceso a contenido digital ofreciendo durabilidad y persistencia.

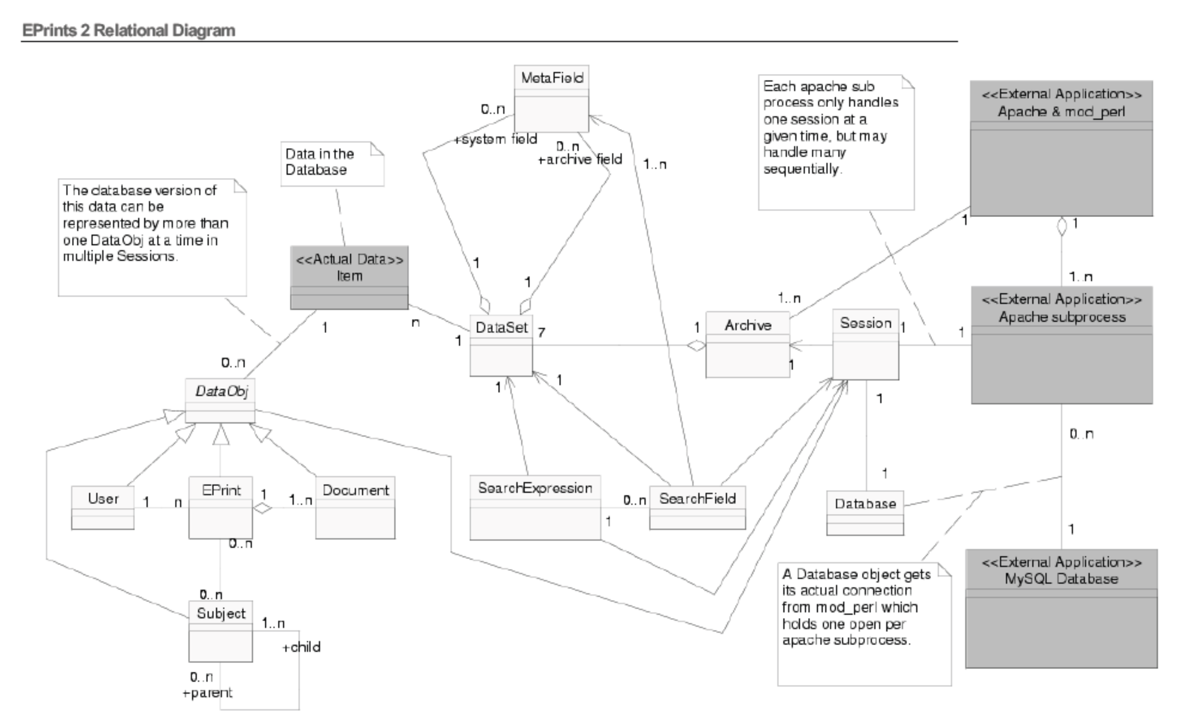
- Amazon S3: Este servicio proporciona una sencilla interfaz de servicios web que puede utilizarse para almacenar y recuperar recursos digitales. Concede acceso a todos los desarrolladores a la misma infraestructura, es altamente escalable, fiable, segura y rápida. Amazon S3 utiliza Amazon para tener en funcionamiento su propia red internacional de sitios web. Este servicio tiene como fin maximizar las ventajas del escalado y trasladar estas ventajas a los desarrolladores.

Figura 11. Estructura General del Núcleo de E-Prints



Fuente: (UNIVERSITY OF SOUTHAMPTON, 2011)

Figura 12. Diagrama Relacional de E-Prints



Fuente: (UNIVERSITY OF SOUTHAMPTON, 2011)

5. FEDORA COMMONS 3.4.2

A continuación se analiza la herramienta para la gestión de repositorios digitales FEDORA Commons versión 3.4.2 (FEDORA, 2005).

5.1 Descripción General

FEDORA (Flexible Extensible Digital Object Repository Architecture) (FEDORA, 2005), es un repositorio flexible de contenido digital y con una gran variedad de usos.

FEDORA se inicia en 1997 en la Universidad de Cornell (CORNELL UNIVERSITY, 2011), como dos proyectos de investigación integrados llamados DARPA y NFS, desarrollados por Sandra Payette, Carl Lagoze y Naomi Dushay. FEDORA se desarrolló a gran escala en el 2002 y en Mayo de 2003 fue liberada la versión 1.0 (FEDORA, 2005).

5.2 Requisitos de software

Los requisitos de software para el correcto funcionamiento de FEDORA (FEDORA, 2005) son:

- Software de desarrollo: Java SE Development Kit (JDK) 6.
- FEDORA utiliza una Base de Datos Relacional para soportar sus funciones por lo que el instalador incluye DerbySQL Database 10.5.3 pero también es compatible con: MySql, Oracle, PostgresSQL y Microsoft.
- Motor Servlet (servidor): Tomcat 6.0.2, Jetty o Jboss
- Compilador: Maven 2

5.3 Facilidad de Personalización

FEDORA implementa el concepto de modelo de objeto digital (FEDORA, 2005). El modelo de objeto digital es un patrón que describe el recurso digital, independiente del formato y de la información que representa. De esta manera, el recurso digital es descrito utilizando el mismo objeto digital sin importar si son datos, definiciones de comportamiento o mecanismos de comportamiento. También es indiferente que tipo de datos está representando el objeto digital: texto, imágenes, video, audio, mapas o datos geoespaciales.

Este mecanismo de representación ofrece las siguientes ventajas para la facilidad de personalización:

- Abstracción: Hace referencia a la facilidad que ofrece FEDORA para gestionar cualquier tipo recurso digital, gracias al modelo de objeto digital utilizado como mecanismo de representación.

- Flexibilidad: FEDORA permite configurar la estructura del modelo de objeto digital para representar los recursos digitales específicos del repositorio.

- Genérico: En FEDORA, los metadatos de descripción y el contenido del recurso digital están empaquetados y fuertemente conectados dentro del objeto digital. FEDORA provee las herramientas y mecanismos para adaptar el conjunto de metadatos de descripción de acuerdo con las características de los recursos digitales del repositorio.

- Agregación: Los objetos digitales gestionados por FEDORA, pueden hacer referencia a información almacenada localmente o en dispositivos remotos. El acceso remoto se provee por medio de herramientas web.

Sin embargo para poder realizar los cambios convenientes en el repositorio es necesario el conocimiento avanzado del lenguaje de programación Java.

Figura 13. Repositorio Digital Islandora soportado por FEDORA Commons



Fuente: (ISLAND, 2011)

5.4 Documentación disponible

FEDORA provee documentación para facilitar su estudio, instalación, configuración y uso (FEDORA, 2005). FEDORA Integra un conjunto de recursos con descripciones introductorias, manuales de usuario de diferentes niveles, tutoriales, herramientas que resuelven preguntas frecuentes y foros de discusión sobre la herramienta. La comunidad de usuarios FEDORA actualiza permanentemente los foros de discusión y desarrolla eventos en diferentes ámbitos geográficos donde los desarrolladores retroalimentan los requisitos de nuevas funcionalidades y cambios para las versiones futuras. Aunque existe una buena cantidad de documentación, se considera que el nivel de detalle no es suficiente para el aprendizaje e implementación.

Figura 14. Documentación de FEDORA



Fuente: (FEDORA, 2005)

5.5 Comunicación con el Usuario

FEDORA provee altos niveles de extensibilidad (FEDORA, 2005). La interfaces de Usuario que ofrece FEDORA son configurables. Esta facilidad se obtiene gracias a que los servicios están directamente asociados con los datos empaquetados en el objeto digital, de esta manera, cuando los servicios cambian, los objetos heredan automáticamente los cambios.

El manejo de la interfaz gráfica en FEDORA se realiza a través de una API llamada API-M.

FEDORA ofrece las siguientes alternativas para el manejo de interfaces de usuario:

- JMS FEDORA es un proveedor JMS (**Java Message Service**). La API **JMS**, es la solución creada por Sun Microsystems para el uso de colas de mensajes. Este es un estándar de mensajería que permite a los componentes de aplicaciones basados en la plataforma Java2, crear, enviar, recibir y leer mensajes. También hace posible la comunicación confiable de manera síncrona y asíncrona enviando mensajes de eventos API-M que administran las operaciones API-M. API-M es un servicio que define una interfaz abierta para administrar el repositorio (FEDORA, 2005).

- JAVA nativo es un framework de programación que permite que un programa escrito en Java ejecutado en la máquina virtual java (JVM) pueda interactuar con programas escritos en otros lenguajes como C, C++ y

ensamblador. Se usa para escribir métodos nativos que permitan solventar situaciones en las que una aplicación no puede ser enteramente escrita en Java, como por ejemplo en el caso de que la biblioteca standard de clases no proporcione soporte para funcionalidades dependientes de la plataforma. Esta alternativa de interfaz de usuario provee una interfaz nativa Java por medio de la API-M. También es posible integrar una interfaz Java simplificada que trace las operaciones API-M ya existentes (FEDORA, 2005).

- Repositorio de contenido Java (JSR 170): Esta alternativa, provee el estándar de interfaz JSR170. JSR-170 es una API tipo JDBC para repositorios de contenido (FEDORA, 2005).

Además de las alternativas de interfaz de usuario propuestas por FEDORA, se pueden implementar propuestas de la comunidad de usuarios. Por ejemplo, el proyecto FEZ es una opción de configuración de interfaces desarrollada en lenguaje PHP.

5.6 Mecanismos de Descripción de recursos digitales

Para la descripción de los recursos digitales del repositorio, FEDORA define dos tipos de metadatos: Administrativos y descriptivos. Los metadatos administrativos los define utilizando la herramienta FOXML.

FOXML es un formato XML simple que expresa directamente el modelo de objetos digitales de FEDORA, por lo tanto, los objetos son almacenados internamente en el repositorio en el formato FOXML. También sirve para importar y exportar objetos de un repositorio FEDORA hacia otro repositorio FEDORA (FEDORA, 2005).

Para poder realizar la gestión de los objetos digitales, estos deben estar registrados con metadatos administrativos. Los metadatos FOXML son requeridos por la arquitectura del repositorio para su correcto funcionamiento. Los metadatos descriptivos son parte del objeto digital y describen sus particularidades de contenido, son opcionales y configurables. Los metadatos descriptivos se utilizan también para vincular objetos digitales entre sí.

5.7 Estadísticas

FEDORA no provee directamente herramientas para el manejo de estadísticas, sin embargo, existen proyectos de interfaces compatibles con FEDORA que implementan opciones para tal fin. Por ejemplo, el proyecto MURADORA

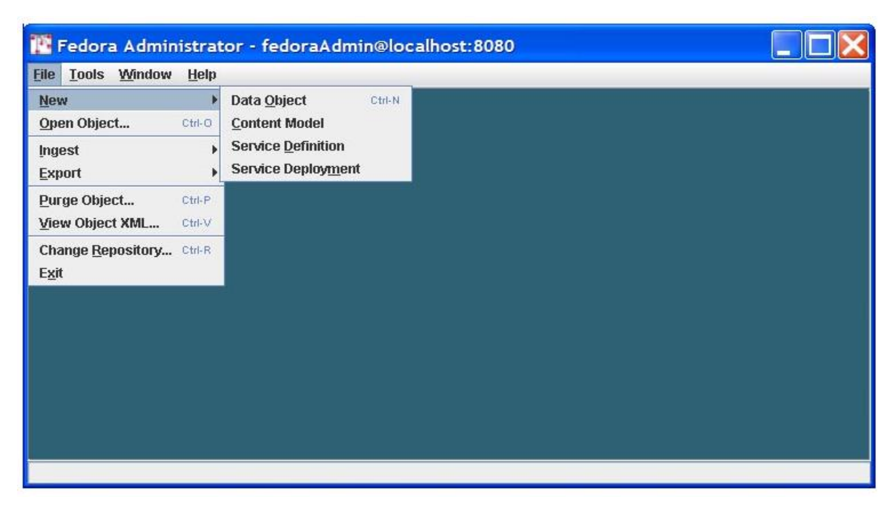
(FEDORA, 2005) desarrolló una aplicación java que recolecta información estadística del acceso y uso del repositorio en una base de datos. Esta base de datos puede ser exportada a otras herramientas para el correspondiente análisis estadístico.

5.8 Gestión de Usuarios

FEDORA utiliza XACML (eXtensible Access Control Markup Language) dentro de su arquitectura de seguridad. XACML está basado en XML y permite codificar las políticas de control de acceso a servicios web. Este control opera tanto a nivel del recurso digital en particular como a nivel del repositorio en general, administrando además los permisos asignados a usuario y grupos de usuarios (FEDORA, 2005).

Los tipos de usuario de FEDORA son: administrador y usuario simple. El administrador principal es el encargado de configurar las políticas de usurario y los recursos digitales del repositorio. FEDORA integra la herramienta FEDORA Administrator (Figura 15), la cual es una conexión directa hacia la funcionalidad API-M en donde se manejan fácilmente las políticas del repositorio sobre cada recurso digital. El usuario simple hace uso del repositorio restringido por el conjunto de políticas de acceso creadas por el administrador.

Figura 15. Herramienta de Administración de FEDORA Commons



Fuente: (FEDORA, 2005)

5.9 Gestión de los recursos digitales

El marco de trabajo de servicios de FEDORA incluye el servicio de repositorio que permite la creación, administración, almacenamiento, acceso y reutilización de los

objetos digitales. Este servicio puede correr en modo stand-alone o situarse dentro del marco de trabajo de servicios de una arquitectura SOA (Service-Oriented Architecture). El núcleo del repositorio es accedido por medio de la interfaz de servicio web (FEDORA, 2005).

5.10 Mecanismos de búsqueda

FEDORA utiliza un conjunto de metadatos configurables basados en el estándar DC (Dublin Core) para la descripción de los recursos del repositorio digital. La herramienta de administración FEDORA permite realizar los procesos de indexación de metadatos y configuración de la búsqueda. La configuración de la búsqueda es una tarea administrativa que consiste en definir los metadatos utilizados como criterios de búsqueda y los metadatos cuyo contenido será mostrado como resultado. FEDORA proporciona dos tipos de búsqueda: Simple y avanzada (FEDORA, 2005).

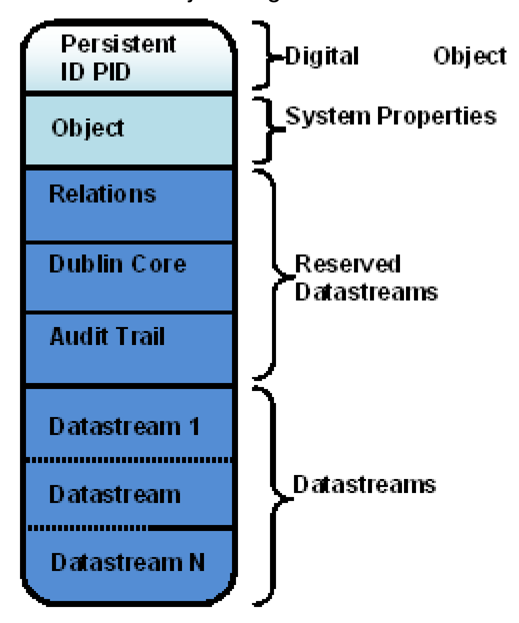
La búsqueda simple consiste en comparar el texto introducido con todos los campos de metadatos indexados para identificar los resultados.

La búsqueda avanzada permite a los usuarios del repositorio desarrollar búsquedas a partir de valores específicos para metadatos específicos.

5.11 Mecanismos de Recuperación

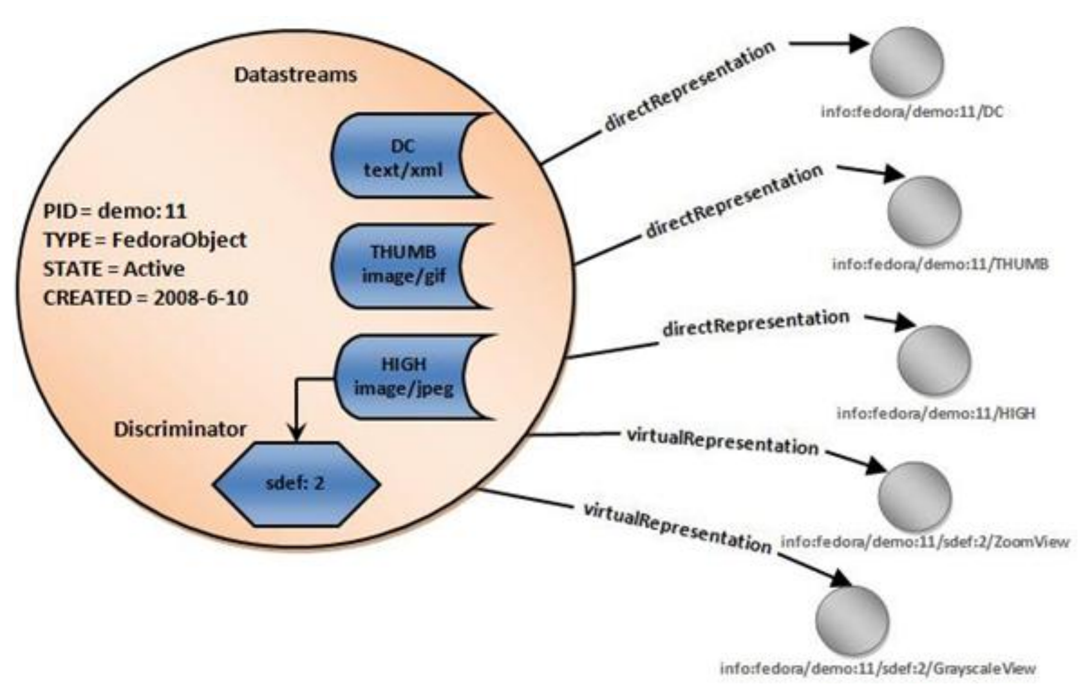
El objeto digital utilizado por FEDORA para representar los recursos contiene el flujo de datos y el conjunto de propiedades del objeto. Hay métodos que definen los puntos de accesos al objeto. Cada punto de acceso permite desplegar una representación del objeto digital. FEDORA provee varias APIs tipo protocolo para acceder a objetos digitales y a su representación (FEDORA, 2005).

Figura 16. Componentes de un Objeto digital en FEDORA



Fuente: (FEDORA, 2005)

Figura 17. Perspectiva de acceso del objeto digital en FEDORA



Fuente: (FEDORA, 2005)

Los puntos de acceso se conectan al marco de trabajo de servicios FEDORA que son invocados por el repositorio para producir representaciones del objeto. Existen puntos de acceso personalizados que permiten producir representaciones directas o virtuales.

5.12 Gestión de experiencias

FEDORA provee facilidades para la personalización de la gestión de experiencias por medio de la edición de código para la creación de herramientas como chats y foros. La personalización de la gestión de experiencias en FEDORA depende de la habilidad de los desarrolladores del repositorio digital (FEDORA, 2005).

5.13 Mecanismos de almacenamiento

Los mecanismos empleados por FEDORA para archivar y preservar los recursos digitales del repositorio son (FEDORA, 2005):

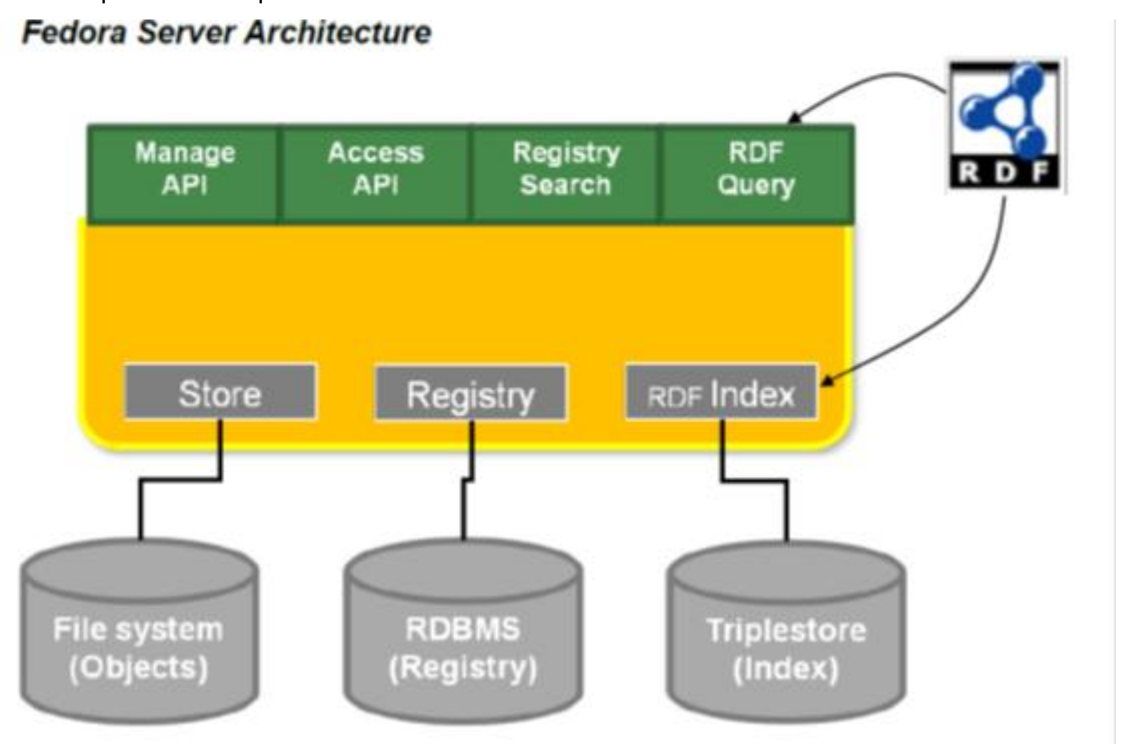
- XML: El contenido XML de los objetos digitales y el esquema sobre el cual están basados son preservados al momento de transferir contenido, durante el almacenamiento y al momento de exportar.

- Versión de contenido: FEDORA ofrece la opción de controlar la versión de los objetos de datos. Cuando un objeto de datos es versionado el rastro de auditoría del objeto es actualizado para reflejar los cambios que se hicieron sobre el objeto, cuando se hicieron los cambios, por quien se hicieron y una nueva versión de los datos modificados es añadida al XML del objeto. Este mecanismo permite regresar a versiones anteriores de objetos de datos por medio de una búsqueda de fecha y tiempo.

- Relaciones objeto a objeto: las relaciones entre objetos pueden ser almacenadas por medio de los metadatos incluidos en los objetos. Esto permite al usuario unir los objetos en una relación padre e hijo.

- Historia de evento: cada objeto en el repositorio FEDORA contiene un rastro de auditoría, en el cual se preserva los registros de los eventos y cambios realizados.

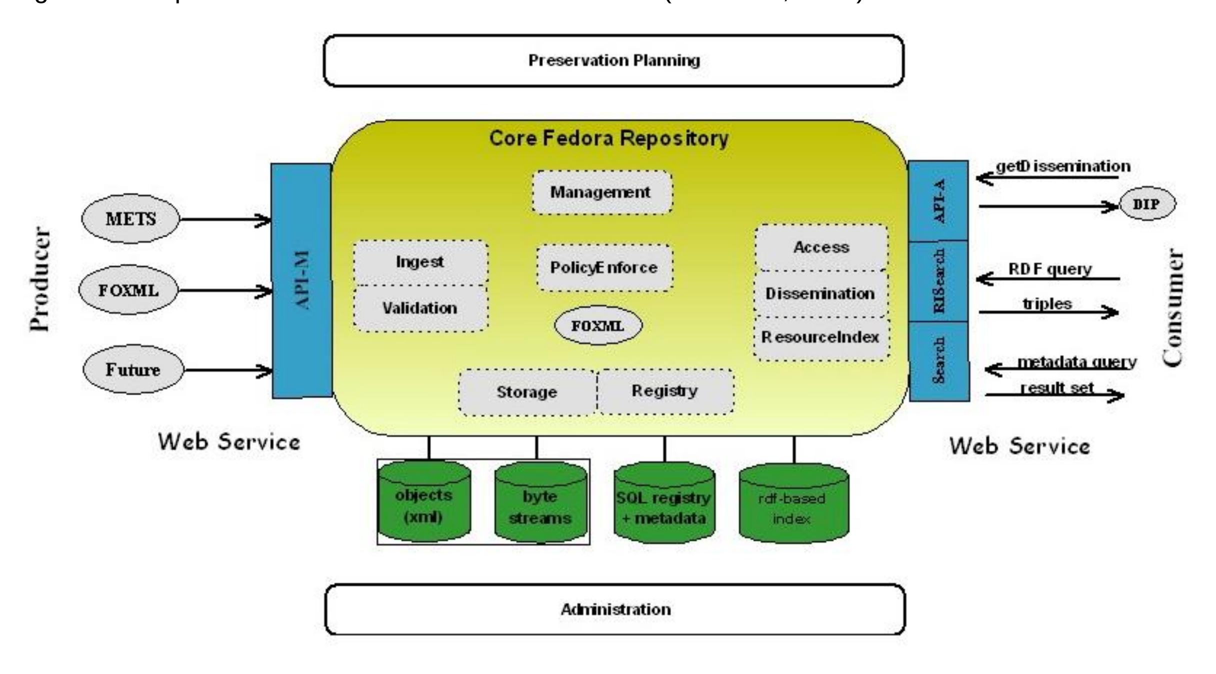
Figura 18. Arquitectura simplificada del sistema FEDORA



Fuente: (FEDORA, 2005)

En la figura 18, se muestra como los usuarios interactúan con el contenido del repositorio por medio de las aplicaciones tipo cliente, visualizadores web, bloques de programa o aplicaciones de servidor. Estas aplicaciones acceden a los datos del repositorio por medio de las cuatro APIs: administración, acceso, búsqueda y la API proveedor OAI la cual está expuesta vía HTTP. Los usuarios pueden interactuar con los objetos únicamente por medio de estas APIs.

Figura 19. Arquitectura del núcleo central de FEDORA (FEDORA, 2005)



Fuente: (FEDORA, 2005)

6. RESUMEN DEL ESTUDIO

A continuación se muestra un resumen del estudio comparativo de herramientas para la gestión de repositorios digitales (tabla 1), se presentan los criterios de comparación más relevantes y las calificaciones asociadas. Las calificaciones fueron asignadas de acuerdo con la experiencia de estudio y prueba de las herramientas comparadas.

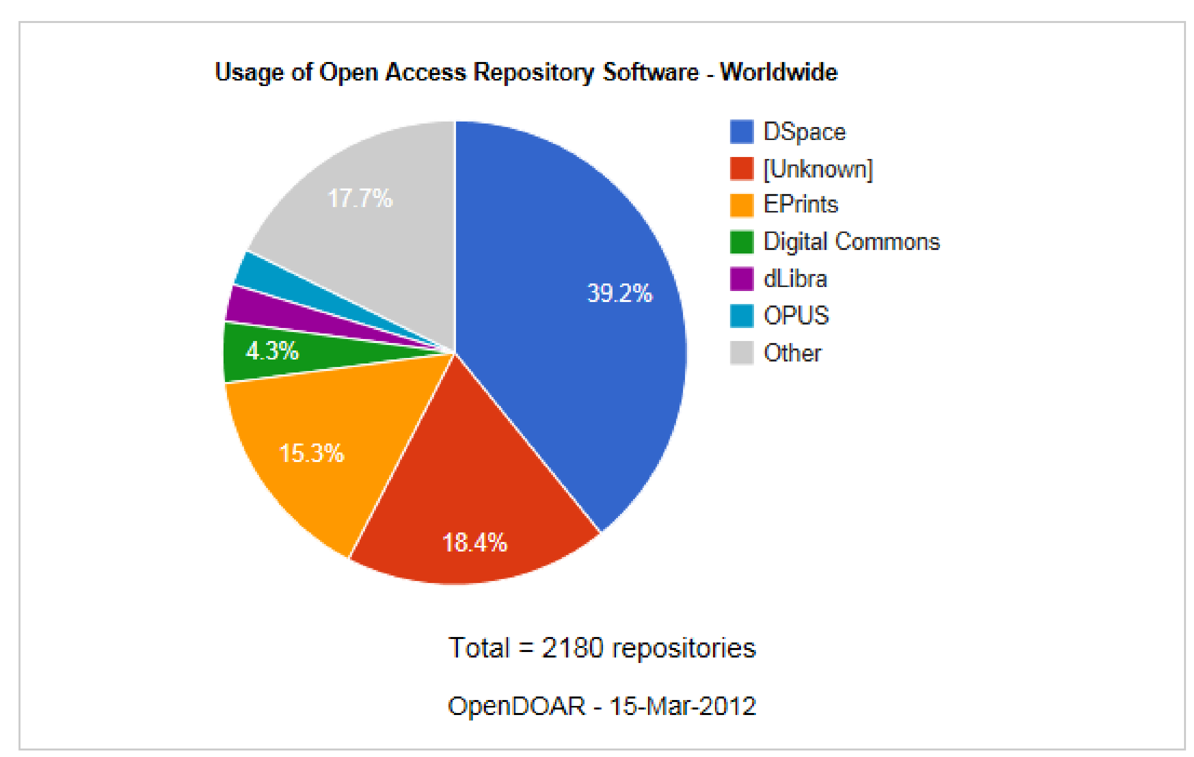
Tabla 1. Resumen de Estudio de las Herramientas para la Gestión de Repositorios Digitales

Criterio	Herramientas		
	DSpace	**E-Prints**	**Fedora Commons**
Facilidad de personalización	Provee potentes herramientas para un nivel de programación básico y avanzado	Provee herramientas para un nivel avanzado de programación	Provee herramientas para un nivel avanzado de programación
Calificación	9,0	9,0	8,0
Comunica-ción con el Usuario	Provee variedad de herramientas de configuración de interfaces para nivel básico y avanzado de programación	Provee herramientas de configuración de interfaces para nivel básico y avanzado de programación	Provee herramientas de configuración de interfaces para nivel básico y avanzado de programación
Calificación	9,0	8,0	8,0
Documentación	Provee varias opciones de fácil acceso para aprendizaje, configuración y uso	Provee documentos de poco detalle. Existen opciones que tienen costo para el usuario	Provee documentos con detalle insuficiente para aprendizaje y configuración
Calificación	9,0	8,0	8,0
Mecanismos de búsqueda	Provee potentes herramientas para configuración de búsqueda básica y avanzada para un nivel básico de programación	Provee herramientas para configuración de búsqueda básica y avanzada para un nivel básico y avanzado de programación	Provee herramientas para configuración de búsqueda básica y avanzada para un nivel avanzado de programación
Calificación	9,0	8,5	8,0
Mecanismos de almacenamiento	Provee un gestor robusto y sofisticado para la persistencia de recursos de forma ilimitada	Provee sofisticados servicios de gestión de persistencia de recursos digitales	Provee un sistema básico de almacenamiento, con opciones de versión de recursos
Calificación	9,0	9,0	8,5
Promedio por Herramienta	**9,0**	**8,5**	**8,1**

7. DSPACE: LA HERRAMIENTA SELECCIONADA

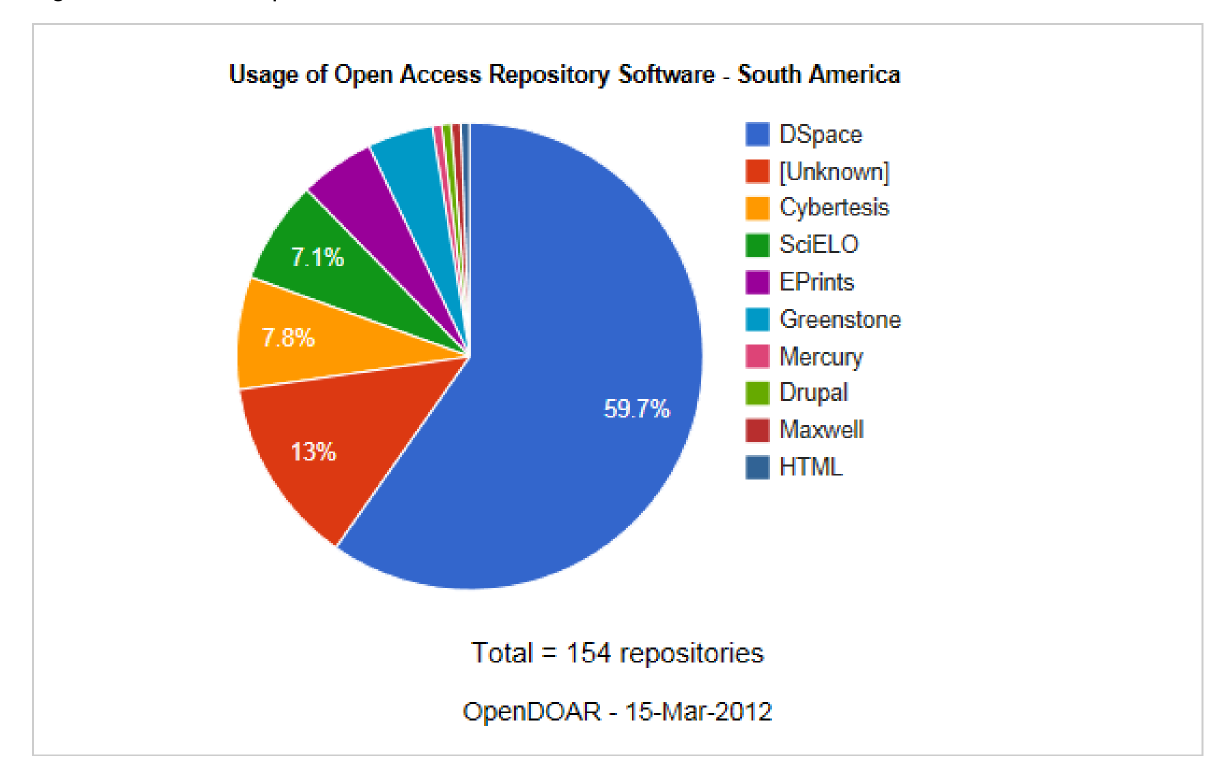
Como se observa en las figuras 13 y 14, OpenDOAR (Directorio autoritativo de repositorios académicos de libre acceso), presenta una serie de estudios estadísticos que concluyen que DSpace es una de las herramientas de gestión de recursos digitales más utilizadas a nivel mundial (UNIVERSITY OF NOTTINGHAM, 2006).

Figura 20. Uso de repositorios de libre acceso a nivel mundial



Fuente: (UNIVERSITY OF NOTTINGHAM, 2006)

Figura 21. Uso de repositorios de libre acceso en Sur América



Fuente: (UNIVERSITY OF NOTTINGHAM, 2006)

DSpace es utilizada por instituciones de investigación a nivel mundial para satisfacer necesidades de archivo digital como: repositorios institucionales, repositorios de objetos de aprendizaje, tesis Electrónicas, administración de registros electrónicos, preservación digital y publicación, entre otras.

Teniendo en cuenta los resultados del estudio comparativo (tabla 1) y los datos estadísticos presentados por OpenDOAR, se ha optado por utilizar DSpace como la herramienta para la implementación de un sistema de gestión de activos de software basado en *PEGASO*.

DSpace proporciona las funcionalidades y herramientas requeridas para implementar el catálogo de activos de software aplicando la propuesta de este trabajo de investigación: facilidades de personalización y comunicación con el usuario, herramientas para la gestión, búsqueda y recuperación de activos, herramientas para la gestión de experiencias y facilidades para la persistencia de los activos y del catálogo.

8. REFERENCIAS

CORNELL UNIVERSITY. 2011. CORNELL. [En línea] 2011. [Citado el: 18 de Septiembre de 2011.] http://www.cornell.edu/.

FEDORA. 2011. Islandora and Discovery Garden, Dspace. [En línea] 2011. [Citado el: 18 de Septiembre de 2011.] https://wiki.duraspace.org/display/DSPACE/Home.

—. 2005. Islandora and Discovery Garden, Fedora Common. [En línea] 2005. [Citado el: 18 de Septiembre de 2011.] http://fedora-commons.org.

HP COMPANY. 2011. HP. *HP.* [En línea] 2011. [Citado el: 18 de Septiembre de 2011.] www.hp.com.

ISLAND, UNIVERSITY OF PRINCIPE EDWARD. 2011. UPEI. *UPEI.* [En línea] UPEI, 2011. [Citado el: 18 de Septiembre de 2011.] http://islandora.ca/.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY. 2011. MIT. *MIT.* [En línea] MIT, 2011. [Citado el: 18 de Septiembre de 2011.] http://web.mit.edu/.

UNIVERSITY OF NOTTINGHAM, U.K. 2006. OPENDOAR. [En línea] 2006. [Citado el: 18 de Septiembre de 2011.] http://www.opendoar.org/index.html.

UNIVERSITY OF SOUTHAMPTON. 2011. E-prints. [En línea] 2011. [Citado el: 18 de Septiembre de 2011.] http://www.eprints.org.

—. 2011. SOTON. *SOTON.* [En línea] 2011. [Citado el: 18 de Septiembre de 2011.] http://www.soton.ac.uk/.